

SciMatics SciQSAR model for the Bacterial Reverse Mutation Test (Ames test) in *S. typhimurium in vitro*

1. QSAR identifier

1.1 QSAR identifier (title)

SciMatics SciQSAR model for the Bacterial Reverse Mutation Test (Ames test) in *S. typhimurium in vitro*, Danish QSAR Group at DTU Food.

1.2 Other related models

MultiCASE CASE Ultra model for the Bacterial Reverse Mutation Test (Ames test) in *S. typhimurium in vitro*, Danish QSAR Group at DTU Food.

Leadscope Enterprise model for the Bacterial Reverse Mutation Test (Ames test) in *S. typhimurium in vitro*, Danish QSAR Group at DTU Food.

1.3. Software coding the model

SciQSAR version 3.1.00.

2. General information

2.1 Date of QMRF

January 2015.

2.2 QMRF author(s) and contact details

QSAR Group at DTU Food;

Danish National Food Institute at the Technical University of Denmark;

<http://qsar.food.dtu.dk/>;

qsar@food.dtu.dk

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Marianne Dybdahl;

National Food Institute at the Technical University of Denmark;

Sine Abildgaard Rosenberg;

National Food Institute at the Technical University of Denmark;

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

Gunde Egeskov Jensen;

National Food Institute at the Technical University of Denmark;

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Danish QSAR Group at DTU Food;

National Food Institute at the Technical University of Denmark;

<http://qsar.food.dtu.dk/>;

qsar@food.dtu.dk

2.6 Date of model development and/or publication

January 2014.

2.7 Reference(s) to main scientific papers and/or software package

Contrera, J.F., Matthews, E.J., Kruhlak, N.L., and Benz, R.D. (2004) Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modelling of the human maximum recommended daily dose. *Regulatory Toxicology and Pharmacology*, 40, 185 – 206.

SciQSAR (2009) Reference guide: *Statistical Analysis and Molecular Descriptors*. Included within the SciMatics SciQSAR software.

2.8 Availability of information about the model

The training set is non-proprietary and was kindly provided by Kazius *et al.* (2005). The model algorithm is proprietary from commercial software.

2.9 Availability of another QMRF for exactly the same mode

3. Defining the endpoint

3.1 Species

Salmonella typhimurium (multiple strains).

3.2 Endpoint

QMRF 4.10. Mutagenicity

OECD 471 Bacterial Reverse Mutation Test

3.3 Comment on endpoint

The bacterial reversed mutation *in vitro* assay using *Salmonella typhimurium* is also referred to as the Ames test. The test is used to evaluate compounds mutagenic properties as it detects point mutations, which involve substitution, addition or deletion of one or a few DNA base pairs. The test uses amino acid-dependent strains of *S. typhimurium*. These strains contain a mutation that makes them unable to synthesize the amino acid histidine. Therefore, in the absence of an external histidine source, the bacteria cannot grow and form colonies. Colony growth is resumed if a reversion of the mutation occurs, allowing the production of histidine to be resumed. Spontaneous reversions occur within each of the strains. If a compound cause an increase in the number of revertant colonies relative to the background level it is said to be positive in the Ames test and therefore assigned mutagenic. Different strains of *S. typhimurium* exist and these have several features that make them more sensitive for the detection of mutations, including responsive DNA sequences at the reversion sites, increased cell permeability to large molecules and elimination of DNA repair systems or enhancement of error-prone DNA repair processes. The specificity of the test strains can provide useful information on the types of point mutations that are induced such as frameshift mutations or base-pair mutations.

Point mutations are the cause of many human genetic diseases and there is substantial evidence that point mutations in oncogenes and tumour suppressor genes of somatic cells are involved in tumour formation in humans and experimental animals. The bacterial reverse mutation test is rapid, inexpensive and relatively easy to perform and for these reasons it has become a useful tool as an initial screen for potential *in vivo* genotoxic activity, and is present the most extensively used *in vitro* short-term test in the screening for point mutation-inducing activity. The test utilises prokaryotic cells, which differ from mammalian cells in such factors as uptake, metabolism, chromosome structure and DNA repair processes. Tests conducted *in vitro* generally require the use of an exogenous source of metabolic activation. *In vitro* metabolic activation systems cannot mimic entirely the mammalian *in vivo* conditions. The test therefore does not provide direct information on the mutagenic and carcinogenic potency of a substance in mammals. It has though been demonstrated that many chemicals that are positive in this test also exhibit mutagenic activity in other tests. For certain classes of chemicals, for example highly bactericidal compounds (e.g. certain antibiotics) and those which are thought (or known) to interfere specifically with the mammalian cell replication system, this test may not be appropriate. Also, there are carcinogens that are not detected by this test because they act through other, non-genotoxic mechanisms or mechanisms absent in bacterial cells.

All chemicals in the training set The categorization of each compound as either a mutagen or a nonmutagen, which was based on the available, occasionally conflicting, Ames test results is described under 3.6. (Kazius *et al.* 2005).

The data used to train this model were compiled by Kazius and co-workers (2005) from multiple sources. All the structures in the data set have experimental results in one or more of the following *S. typhimurium* tester strains: TA98, TA100, TA1535 and either TA1537 or TA97. Strains TA102 and TA1538 were also applied in cases where results of other strains are equivocal or difficult to interpret. The inclusion criteria for the data as well as the categorization of chemicals in to Ames mutagens or non-mutagens is described in Kazius *et al.* (2005): "...Ames tests were only considered if they were performed with the standard plate method or the preincubation method, either with or without a metabolic activation mixture. Second, this study required the categorization of each compound as either a mutagen or a nonmutagen, which was based on the available, occasionally conflicting, Ames test results determined in different laboratories. In this study, a compound was categorized as a mutagen if at least one Ames test result was positive. Consequently, a false positive Ames test result will erroneously rendering a compound mutagenic, irrespective of the number of negative results. In general, the categorization of a compound as nonmutagenic is sufficiently reliable if at least four Ames tests, performed with different strains, give reproducible negative results. In this study, to assemble a large dataset with maximal compound diversity, a compound was categorized as a nonmutagen if exclusively negative Ames test results - one or more - were reported. Further, the robustness of the above mutagenicity categorization of the CCRIS database was tested by applying the same categorization criteria to another set of Ames test results collected from the NTP (National Toxicology Program). The results obtained for approximately 1500 compounds present in both the NTP and the CCRIS databases showed contradicting categorizations in 11% of the cases. Because this error was smaller than 15%, which is the average interlaboratory reproducibility error of Ames tests, the categorization applied in this study was considered satisfactory. To further increase the consistency of the dataset, compounds whose CCRIS data showed contradicting categorizations with the NTP data were removed from the dataset. In conclusion, a dataset of 4337 compounds with corresponding molecular structures and toxicity categorizations (2401 mutagens and 1936 nonmutagens) was constructed."

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Mutagenic in the Bacterial Reverse Mutation Test (Ames test) in *Salmonella Typhimurium in vitro*, positive or negative.

3.6 Experimental protocol

The experimental protocol is described in OECD guideline 471 (1997). Briefly, suspensions of bacterial cells are exposed to the test substance in the presence and in the absence of an exogenous metabolic activation system. The most commonly used system is a cofactor supplemented post-mitochondrial fraction (called S9) prepared from the livers of rodents. In the plate incorporation method, these suspensions are mixed

with an overlay agar and plated immediately onto minimal medium. In the preincubation method, the treatment mixture is incubated and then mixed with an overlay agar before plating onto minimal medium. For both techniques, after two or three days of incubation, revertant colonies are counted and compared to the number of spontaneous revertant colonies on solvent control plates (OECD guideline 471, 1997).

3.7 Endpoint data quality and variability

As data originates from multiple sources and consist of a combination of results from different *S. typhimurium* tester strains some degree of variability in the data is expected. Further, as described by Kazius *et al.* (2005): "The reproducibility of Ames tests is limited by the purity of the tested chemical, inconsistencies in the interpretation of dose-response curves, interference of further toxic side effects (such as cytotoxicity), variations in the methodology employed, and variations in the materials used (bacterial strains and metabolic activation mixtures). Nevertheless, the average interlaboratory reproducibility of a series of Ames test data from the National Toxicology Program (NTP) was determined to be 85%."

4. Defining the algorithm

4.1 Type of model

This is a categorical (Q)SAR model based on calculated molecular descriptors, and if available the modeller's own or third-party descriptors or measured endpoints can be imported and used as descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model made by use of parametric discriminant analysis to create a linear discriminant function (see 4.5). The specific implementation is proprietary within the SciQSAR software.

4.3 Descriptors in the model

Molecular connectivity indices

Molecular shape indices

Topological indices

Electrotopological (Atom E and HE-States) indices

Electrotopological bond types indices

SciQSAR software provides over 400 built-in molecular descriptors. Additionally, SciQSAR makes it possible to import the modeller's own or third-party descriptors or use measured endpoints as custom descriptors.

4.4 Descriptor selection

The initial descriptor set is manually chosen by the model developer from the total set of built-in descriptors. Furthermore, the set of descriptors applied in the modelling by the program is on top of this selection determined by thresholds for descriptor variance and number of nonzero values likewise defined by the model developer.

151 descriptors were selected from the initial pool of descriptors by the system and used to build the model.

4.5 Algorithm and descriptor generation

For a binary classification problem SciQSAR uses discriminant analysis (DA) to make a (Q)SAR model. SciQSAR implements a broad range of discriminant analysis (DA) methods including parametric and non-

parametric approaches. The classic parametric method of DA is applicable in the case of approximately normal within-class distributions. The method generates either a linear discriminant function (the within-class covariance matrices are assumed to be equal) or a quadratic discriminant function (the within-class covariance matrices are assumed to be unequal). When the distribution is assumed to not follow a particular law or is assumed to be other than the multivariate normal distribution, non-parametric DA methods can be used to derive classification criteria. The non-parametric DA methods available within SciQSAR include the kernel and *k*-nearest-neighbor (kNN) methods. The main types of kernels implemented in SciQSAR include uniform, normal, Epanechnikov, bi-weight, or tri-weight kernels, which are used to estimate the group specific density at each observation. Either Mahalanobis or Euclidean distances can be used to determine proximity between compound-vectors in multidimensional descriptor space. When the kNN method is used, the Mahalanobis distances are based on the pooled covariance matrix. When the kernel method is used, the Mahalanobis distances are based on either the individual within-group covariance matrices or the pooled covariance matrix. (Contrera *et al.* 2004)

If the data outcome is continuous, regression analysis is used to build the predictive model. Within SciQSAR several regression methods are available: ordinary multiple regression (OMR), stepwise regression (SWR), all possible subsets regression (PSR), regression on principal components (PCR) and partial least squares regression (PLS). The choice of regression method depends on the number of independent variables and whether correlation or multicollinearity among the independent variables exists: OMR is acceptable with a small number of independent variables, which are not strongly correlated. SWR is used under the same circumstances as OMR but with greater number of variables. PSR is used for problems with a great number of independent variables. PCR and PLS are useful when a high correlation or multicollinearity exist among the independent variables. (SciQSAR 2009)

To test how stable the developed models are, SciQSAR have built-in cross-validation procedures (see 6.).

For this model, the linear method was used.

4.6 Software name and version for descriptor generation

SciQSAR version 3.1.00.

4.7 Descriptors/chemicals ratio

In this model 151 descriptors were used. The training set consists of 4,102 compounds. The descriptor/chemical ratio is 1:27.2 (151:4102).

5. Defining Applicability Domain

5.1 Description of the applicability domain of the model

The definition of the applicability domain consists of two components; the definition in SciQSAR and the in-house further refinement algorithm on the output from SciQSAR to reach the final applicability domain call.

1. SciQSAR

The first criterion for a prediction to be within the models applicability domain is that all of the descriptor values for the test compound can be calculated by SciQSAR. If SciQSAR cannot calculate each descriptor value for the test chemical no prediction value is given by SciQSAR and it is considered outside the model's applicability domain.

2. The Danish QSAR group

The Danish QSAR group has applied a stricter definition of applicability domain for its SciQSAR models. In addition to the applicability domain definition made by SciQSAR a second criterion has been applied for predictions generated from (Q)SAR models with a binary endpoint. For each prediction SciQSAR calculates the probability (p) for the test compound's membership in one of the two outcome classes (positive or negative). The probability of membership in a class is a measure of how well training set knowledge is able to discriminate a positive prediction from a negative prediction within the nearest space of the subject compound-vector. The probability of membership value is also a measure of the degree of confidence of a prediction. The Danish QSAR group uses this probability for a prediction to further define the model's applicability domain. Only positive predictions with a probability equal to or greater than 0.7 and negative predictions with a probability equal to or less than 0.3 are accepted. Positive predictions with a probability between 0.5 and 0.7 as well as negative predictions with a probability between 0.3 and 0.5 are considered outside the model's applicability domain. When these predictions are wed out the accuracy of the model in general increases at the expense of reduced model coverage. Furthermore, as SciQSAR does not define a structural domain, only predictions which were within either Leadscope structural domain (defined as at least one training set chemical within a Tanimoto distance of 0.7) or CASE Ultra structural domain (no unknown fragments for negatives and maximum 1 unknown fragment for positives) were defined as being inside the SciQSAR applicability domain.

5.2 Method used to assess the applicability domain

The system does not generate predictions if it cannot calculate each descriptor value for the test compound.

Only positive predictions with probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 were accepted.

5.3 Software name and version for applicability domain assessment

SciQSAR version 3.1.00.

5.4 Limits of applicability

The Danish QSAR group applies an overall definition of structures acceptable for QSAR processing which is applicable for all the in-house QSAR software, i.e. not only SciQSAR. According to this definition accepted structures are organic substances with an unambiguous structure, i.e. so-called discrete organics defined as: organic compounds with a defined two dimensional (2D) structure containing at least two carbon atoms, only certain atoms (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I), and not mixtures with two or more 'big components' when analyzed for ionic bonds (for a number of small known organic ions assumed not to affect toxicity the 'parent molecule' is accepted). Structures with less than two carbon atoms or containing atoms not in the list above (e.g. heavy metals) are rendered out as not acceptable for further QSAR processing. Calculation 2D structures (SMILES and/or SDF) are generated by stripping off accepted organic and inorganic ions. Thus, all the training set and prediction set chemicals are used in their non-ionized form. See 5.1 for further applicability domain definition.

6. Internal validation

6.1 Availability of the training set

Yes

6.2 Available information for the training set

CAS

SMILES

6.3 Data for each descriptor variable for the training set

No

6.4 Data for the dependent variable for the training set

All

6.5 Other information about the training set

4102 compounds are in the training set: 2299 positives and 1803 negatives.

6.6 Pre-processing of data before modelling

The original data set from Kazius *et al.* (2005) consisted of 4337 molecular structures with corresponding Ames test data. Of these 235 were excluded in the pre-processing due to:

- Only structures acceptable for the commercial software could be processed
- Only discrete organic chemicals as described in 5.4 were used
- In case of replicate structures, one of the replicates was kept if all the replicates had the same activity and all were removed if they had different activity

6.7 Statistics for goodness-of-fit

SciQSARs own internal performance test of the model gave the following Cooper's statistics for predictions within the applicability domain as defined by SciQSAR (i.e. the first criterion described in 5.1):

- Sensitivity (true positives / (true positives + false negatives)): 80.7%
- Specificity (true negatives / (true negatives + false positives)): 81.4%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 81.1%

6.8 Robustness – Statistics obtained by leave-one-out cross-validation

Not performed.

6.9 Robustness – Statistics obtained by leave-many-out cross-validation

SciQSAR's own internal 10-fold cross-validation (10*10% out) procedure was used for predictions within the applicability domain as defined by SciQSAR (i.e. the first criterion described in 5.1). As the probability domain was not applied (i.e. the second criterion described in 5.2) the accuracy of the predictions when applying this domain can be expected to be higher than reflected in these cross-validation results. This gave the following Cooper's statistics:

- Sensitivity (true positives / (true positives + false negatives)): 79.3%
- Specificity (true negatives / (true negatives + false positives)): 79.1%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 79.2%

6.10 Robustness - Statistics obtained by Y-scrambling

Not performed.

6.11 Robustness - Statistics obtained by bootstrap

Not performed.

6.12 Robustness - Statistics obtained by other methods

Not performed.

7. External validation

7.1 Availability of the external validation set

7.2 Available information for the external validation set

7.3 Data for each descriptor variable for the external validation set

7.4 Data for the dependent variable for the external validation set

7.5 Other information about the training set

7.6 Experimental design of test set

7.7 Predictivity – Statistics obtained by external validation

7.8 Predictivity – Assessment of the external validation set

7.9 Comments on the external validation of the model

External validation not performed.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The SciQSAR software provides over 400 calculated physico–chemical, electrotopological E-state, connectivity and other molecular descriptors. The descriptors selected for the model may indicate modes of action that are obvious for persons with expert knowledge about the endpoint.

8.2 A priori or posteriori mechanistic interpretation

A posteriori mechanistic interpretation. The descriptors selected for the model may provide a basis for mechanistic interpretation.

8.3 Other information about the mechanistic interpretation

9. Miscellaneous information

9.1 Comments

This model can be used to predict if a chemical is an Ames mutagen or non-mutagen according to the categorization made by Kazius and co-workers (2005).

9.2 Bibliography

Kazius, J., McGuire, R., and Burs, R. (2005) Derivation and Validation of Toxicophores for Mutagenicity Prediction. *J. Med. Chem.*, 48, 312-320.

OECD guideline 471 (1997) OECD Guidelines for the Testing of Chemicals No. 471, Bacterial Reverse Mutation Test. Organisation for Economic Cooperation and Development; Paris, France. Available online at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

9.3 Supporting information